PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12N 15/00, 5/00, C07K 7/00 C07K 13/00

A1

(11) International Publication Number: WO 88/ 02402

(43) International Publication Date: 7 April 1988 (07.04.88)

(21) International Application Number: PCT/US87/02401

(22) International Filing Date: 22 September 1987 (22.09.87)

(31) Priority Application Number: 912,408

(32) Priority Date: 26 September 1986 (26.09.86)

(33) Priority Country: US

(71) Applicant: CALGENE, INC. [US/US]; 1920 Fifth Street, Suite F, Davis, CA 95616 (US).

(72) Inventors: COMAI, Luca ; 502 Lake Boulevard, Davis, CA 95616 (US). STALKER, David, M. ; 3317 Oyster Bay Avenue, Davis, CA 95616 (US).

(74) Agent: ROWLAND, Bertram, I.; Leydig, Voit & Mayer, 350 Cambridge Avenue, Suite 200, Palo Alto, CA 94306 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published

*With international search report.*

*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: INTRACELLULAR DIRECTED DELIVERY OF EXPRESSION PRODUCTS

(57) Abstract

Novel chimeric constructions are provided for expression of chimeric genes having an N-terminus transit peptide for efficient translocation from the cytoplasm to the chloroplast and a foreign gene capable of functioning in association with the chloroplast. Particularly, the transit peptide from the small subunit of ribulose-1,5-bisphosphate carboxylase is joined to an *aro*A gene expressing an inhibitor resistant enzyme with retention of the processing signal to provide the mature *aro*A expression product in the chloroplast.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | |
|---|---|
| AT | Austria |
| AU | Australia |
| BB | Barbados |
| BE | Belgium |
| BG | Bulgaria |
| BJ | Benin |
| BR | Brazil |
| CF | Central African Republic |
| CG | Congo |
| CH | Switzerland |
| CM | Cameroon |
| DE | Germany, Federal Republic of |
| DK | Denmark |
| FI | Finland |
| FR | France |
| GA | Gabon |
| GB | United Kingdom |
| HU | Hungary |
| IT | Italy |
| JP | Japan |
| KP | Democratic People's Republic of Korea |
| KR | Republic of Korea |
| LI | Liechtenstein |
| LK | Sri Lanka |
| LU | Luxembourg |
| MC | Monaco |
| MG | Madagascar |
| ML | Mali |
| MR | Mauritania |
| MW | Malawi |
| NL | Netherlands |
| NO | Norway |
| RO | Romania |
| SD | Sudan |
| SE | Sweden |
| SN | Senegal |
| SU | Soviet Union |
| TD | Chad |
| TG | Togo |
| US | United States of America |

# INTRACELLULAR DIRECTED DELIVERY OF EXPRESSION PRODUCTS

## CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation-in-part of Application Serial No. 912,408, filed September 26, 1986.

## INTRODUCTION

### Technical Field

This invention relates to techniques for directing the transport of proteinaceous gene products to particular organelles of a cell in which the product is being produced.

### Background

Plant cells contain distinct subcellular compartments delimited by membranes. In photosynthetic leaf plants, the most conspicuous organelles are the chloroplasts. The chloroplast present in leaf cells is one developmental stage of this organelle. Proplastids, etioplasts, amyloplasts, and chromoplasts are different stages. The embodiments of this invention apply to the organelle "at large," which will be referred to as "chloroplast." The majority of chloroplast proteins are coded by nuclear genes synthesized in the cytoplasm and then imported into the chloroplasts. Import is associated with the removal of an amino terminal portion, the transit peptide. The transit peptide is linked to the mature peptide by an amino acid sequence, normally requiring at least two amino acids, which is recognized by a specific protease associated with the chloroplast. Thus, the proform of the mature peptide is translocated to the chloroplast

and processed as a result of recognition by one or more proteins.

For many purposes in the manipulation and transformation of plant cells to provide particular functions in the plant cell, it will be desirable that the gene which is introduced into the plant cell results in a product which is translocated to the chloroplast and functions in the chloroplast. It is therefore desirable to be able to devise chimeric constructions which provide for efficient expression in a plant host, resulting in the enhanced production of a particular product or the production of a novel product which is translocated to and functions in the chloroplasts of the plant host.

Relevant Literature

Van den Broeck et al., Nature (1985) 313: 358-363 describes the use of the transit peptide of the pea small subunit ribulose-1,5-bisphosphate carboxylase joined to the NPT-II gene to provide a chimeric gene which is shown to be capable of translocation in a plant host. Lubben and Keegstra, Proc. Natl. Acad. Sci. USA (1986) 83:5502-5506 report the use of the soybean SSU transit peptide plus 13 amino acids of the pea SSU mature protein plus a heat shock protein for translocation to chloroplasts. See also the references cited therein as well as Cashmore et al., Biotechnology (1985) 803-808, and Karlin-Neumann et al., EMBO J. (1986) Vol 5, No. 1, pp. 9-13. The mutated aroA gene is described in U.S. Patent No. 4,535,060 and in Comai et al., Science (1983) 221:370, each of which is incorporated herein by reference. See also EPA 218,571.

## SUMMARY OF THE INVENTION

DNA constructs and the resulting proteinaceous products expressed from such constructs are provided

resulting in the translocation of proteins expressed in the cytoplasm into chloroplasts. The constructs involve sequences encoding a transit peptide recognized by the host plant cell for transporting a cytoplasmic expressed protein to the chloroplast, a short polar or hydrophilic region having at least two positively charged amino acids, and a protein of interest which is fused to the short polar region. The polar region may be naturally occurring or synthetic, and serves to provide for efficient transport from the cytoplasm to the chloroplast, usually with efficient processing and removal of the transit peptide.

## DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Methods and compositions are provided for introducing genes, heterologous or homologous, into a plant cell, whereby the gene is expressed and the resulting polypeptide product transported from the cytoplasm to the chloroplast. The gene constructs involve in the direction of transcription of the sense or positive polarity strand, a transcriptional initiation region recognized by the plant host, a translated region, comprising the gene of interest, and a transcriptional termination region recognized by the plant host cell. The translated region comprises a transit peptide recognized by the host for transporting a polypeptide from the cytoplasm into the chloroplast, usually a processing signal which may be the same or different from the processing signal naturally associated with the transit peptide for cleaving the transit peptide upon transport into the chloroplast, a sequence encoding a polar amino acid region having at least two amino acids having positive charges, and the gene of interest, where the transit peptide, processing sequence, polar region sequence, and gene of interest are all in proper reading frame.

A significant number of biological processes are involved in the chloroplast. Therefore, there is substantial interest in having a wide variety of proteins introduced into the chloroplast, where the proteins may result in modification of a biological process, imparting new capabilities to the chloroplast, or protecting a biological process from interference.

Since the chloroplast is the site for production of fatty acids, by introducing various proteins into the chloroplast, one may enhance the production of fatty acids, modify the size distribution of fatty acids or modify the unsaturated character of the fatty acids, both as to number and site. Various enzymes which may be involved in such a function include acyl carrier protein (ACP), acetyl-CoA ACP transacylase, thioesterase, malonyl-CoA ACP transacylase, β-ketoacyl ACP, synthetase, etc.

Another biological process associated with the chloroplast is starch synthesis. This process involves the enzymes starch phosphorylase, NDPG transglycosylase, the "D-enzyme" (Peat et al., Nature (1953) 72:158, the "Q-enzyme" (Baum and Gilbert, Nature (1953) 172:983), ADP glucose pypophosphorylase, etc.

Other proteins which are produced in the cytoplasm and translocated to the chloroplast include the small subunit of the ribulose-1,5-bisphosphate carboxylase, chlorophyll A and B binding protein, ferredoxin, enzymes of the shikimic acid pathway, as well as other amino acid biosynthetic enzymes.

Of particular interest is the biological process for the synthesis of aromatic amino acids which includes in the pathway the enzyme 5-enolpyruvyl-3-phosphoshikimate synthetase. This enzyme is of particular interest because it is the target for a commonly used herbicide glyphosate.

The expression product is a precursor or pro-form of the desired peptide, having a leader sequence recognized by the host for translocation from the cyto-plasm joined to the gene of interest by a peptide sequence recognized as a processing signal. The pro-form of the peptide produced in the cytoplasm is trans-located to the chloroplast, where it is processed by cleavage of a peptide. The processing signal DNA sequence encodes an oligopeptide recognized by a peptidase. Upon transport through the membrane of the chloroplast and entry in the chloroplast organelle, the expression product is cleaved by the peptidase to fulfill its intended function as a mature protein.

Any gene of interest may be joined to a tran-sit peptide encoding sequence and processing signal through a linking sequence encoding a polar region. Any of the genes encoding the proteins indicated above may be employed. As indicated, these genes may be involved in modifying the composition of a natural com-position, providing protection from stress, e.g., her-bicides, enhancing a particular function or the like.

The aroA gene is involved in expression of enzymes in the shikimic acid metabolic pathway, partic-ularly the enzyme that catalyzes the conversion of phosphoenolpyruvate and 5-phosphoshikimic acid to 5-enolpyruvyl-3-phosphoshikimic acid. The enzyme is 5-enolpyruvyl-3-phosphoshikimate synthetase (EC:2.5.1.19); referred to hereafter as ES-3-P synthe-tase. The natural gene product is useful in enhancing the production of shikimic acid in metabolic pathway products. Genes that express an enzyme which is resis-tant to glyphosate (N-phosophonomethyl glycine), an important and commonly used herbicide, are useful in imparting glyphosate resistance to a normally glyphos-ate-sensitive cell in which the structural gene is expressed. U.S. Patent No. 4,535,060 describes such a mutated structural gene.

The transit peptide and processing signal may be derived from any plant protein which is expressed in the cytoplasm and translocated to the chloroplast. By comparing the messenger RNA for the particular polypeptide with the mature product, the amino acid sequence absent from the mature protein and coded for by the messenger beginning at the initiation codon, usually a methionine, will normally be the transit sequence. In many situations, it may be desirable to employ one transit sequence as distinguished from another. Exemplary of transit sequences is the transit peptide of the small subunit (SSU) of the ribulose-1,5-bisphosphate carboxylase and that from acyl carrier protein (ACP). Fragments from the natural transit sequence that retain their transport activity can also be used. The transit peptide is a sequence capable of translocating a peptide joined to the transit peptide to the chloroplast and may be the whole wild-type transit peptide, a functional fragment therof, or a functional mutant thereof. The full natural sequence is naturally also included. For the most part, the transit peptides from one plant are generally recognized by other plants. Thus, the transit peptide may be native to or heterologous to the ultimate host in which the chimeric gene is introduced. Transit peptides may come from soybean, corn, petunias, tobacco, brassica, tomato, wheat, pea, etc. The transit peptide will usually have at least about 20 amino acids and not more than about 100 amino acids.

The translated region will provide for a chimeric gene having three components: (1) the transit peptide and normally the processing signal, which provides for translocation of the expression product from the cytoplasm to the chloroplast and processing for removal of the transit peptide; (2) a polar region which affects the efficiency of the translocation and stability of the gene of interest; and (3) the gene of

interest which provides for a functional protein product which will provide a change in the functioning of the chloroplast.

The polar region which will separate the transit peptide and processing signal from the gene of interest will normally involve at least 6, usually 8, codons and not more than about 20 codons, although there may be specific exceptions. The polar region may be the naturally occurring region of the mature peptide naturally joined to the transit peptide, may be a region obtained from a different protein naturally associated with a transit peptide and providing the N-terminus of such mature protein, or may be a synthetic sequence encoding an unnatural polypeptide sequence.

Besides the number of codons, the sequence will be further characterized by being hydrophilic and desirably having at least about 40 number percent, preferably at least about 50 number percent of the polar amino acids. The polar amino acids comprise the charged and neutral amino acids having a heteroatom in the chain. These amino acids include the charged amino acids, lysine (K), arginine (R), aspartic acid (D), glutamic acid (E), and histidine (H) and the neutral polar amino acids serine (S), threonine (T), asparagine (N), glutamine (Q). Normally, the sequence will include at least two positively charged amino acids, preferably at least three, and may include up to five positively charged amino acids, but usually not more than 25 number percent, preferably not more than 20 number percent of the sequence will be positively charged amino acids. Charged amino acids will be present in the range from 6 to 15 amino acids and may also be present as the first amino acid after the processing signal. Also, with acp, the polar region may be from 12 to 30 amino acids of the N-terminus of the mature acp. To define hydrophilic region, the hydrophilic or hydrophobic character of a protein

region may be determined as described by Kyte and Doolittle, J. Mol. Bio. (1982) 157:105-132. This reference indicates the exposed nature of the region.

As an individual exception, the N-terminal region of the plant acyl carrier protein may be employed, with up to and including 35 codons, preferably not more than about 30 codons and may range from about 12 codons to 30 codons. The polar region will be designed to usually be followed by a hydrophobic region, which will have at least about 50% hydrophobic amino acids, preferably at least about 60% hydrophobic amino acids in the amino acids immediately following the last positively charged amino acid. The particular hydrophobic amino acids include glycine (G), alanine (A), proline (P), and more particularly leucine (L), isoleucine (I), valine (V), and the aromatic amino acids phenylalanine (F), tyrosine (Y), and tryptophan (W). Desirably, the next five amino acids, more desirably the next ten amino acids following the last positively charged amino acid of the polar region will be free of positively charged amino acids. This hydrophobic region can be achieved by employing the polar region, having naturally present the appropriate amino acids as part of the gene of interest, the wild-type sequence associated with the transit peptide, synthetic sequences or combinations thereof.

Desirably, one would have as few amino acids as may be employed to obtain the desired efficiency of translocation. Since the final product will be a fused protein comprising the polar region and gene of interest, the polar region may have an effect on the gene of interest, as to its stability, activity, location in the chloroplast, or other characteristic. Therefore, for the most part, the polar region will be selected, so as to minimize the adverse effect on the role of the gene of interest in the chloroplast.

Depending upon the nature of the polar region, various strategies may be employed for synthesizing the chimeric gene. Where the naturally occurring N-terminal codons of the mature protein normally associated with the transit peptide are employed, the gene of interest will be joined to the polar region in the appropriate reading frame. Where there is a convenient restriction site at the end of the polar region, the transit peptide containing gene may be cleaved at that site and manipulated for joining to the gene of interest. Alternatively, where there is no convenient restriction site, an alternative restriction site may be employed and an adapter used to restore the lost codons. In some instances, resection may be employed with Bal31 to provide for a blunt terminus which may be joined to a blunt ended gene of interest.

The particular manner in which the sequences are joined and ligated is not critical to this invention. Where the polar region is not naturally joined to the transit peptide, it may by synthesized or obtained from any convenient source and be used as an adapter or bridge to link the transit peptide to the gene of interest. In the absence of convenient restriction sites in the transit peptide and processing signal portion and/or the gene of interest, the polar region may serve as an adapter restoring any codons lost by restricting internal to the other regions.

The DNA sequence encoding the transit peptide may be the complete transit-peptide-encoding sequence including the processing signal or a truncated transit-peptide-encoding sequence lacking from about 1 to 10 codons or a portion of a codon from the 3'-terminus. In addition, one or more changes may be made in the nature of mutations, deletions or insertions in the transit peptide and processing signal, where such change may provide for convenience in construction, providing for a convenient restriction site or removing

an inconvenient restriction site. The mutations may be conservative or non-conservative, so that the transit peptide may be the same or different from the wild-type transit peptide.

The individual sequences may be obtained from naturally occuring sources, may be sequences modified from naturally occurring sources, may be combinations of sequences from naturally occurring and synthetic sequences, and the like. Various techniques can be used for modifying these sequences such as in vitro mutagenesis, primer repair, resection, ligation, tailing, etc. The various techniques may be carried out in accordance with conventional procedures.

The structural genes of interest may be derived from cDNA, from chromosomal DNA or may be synthesized in whole or in part. All or a portion of the structural gene may be a naturally occurring sequence coding for a wild-type peptide or a synthetic sequence in whole or in part coding for a wild-type peptide or a mutant, as a result of a point mutation, insertions or deletions. In some situations it may be desirable to modify all or a portion of the codons, for example to enhance expression employing host-preferred codons. Of particular interest are heterologous genes from other than plant sources, such as microorganisms, e.g. bacteria and fungi, non-vertebrates, e.g. insects, and vertebrates, e.g. mammals and fish.

The various segments of the chimeric gene can be joined in a variety of ways, usually involving one or more of the fragments being inserted into a vector, cloned, the cloned plasmid restricted, followed by manipulations, as appropriate, and joining of the adjacent fragment to the cloned fragment. These steps can be repeated until the entire chimeric gene has been completed.

A particularly preferred chimeric gene of the present invention is formed by connecting the transit peptide from the small subunit (SSU) of ribulose 1,5-bisphosphate carboxylase with a mutated aroA gene capable of imparting resistance to glyphosate. Although a number of publications have tended to indicate that the SSU transit peptide would be generally useful without regard to how it is attached to the structural gene, this has proven not to be true when the SSU transit peptide gene is joined to the aroA gene, as well as other genes.

Of particular interest is employing from 10 to 20 codons of the SSU 5'-terminus of the sequence encoding the mature SSU protein and not more than 20 codons, preferably from 14 to 18 codons, more preferably 16 codons.

For example, Van den Broeck et al., Nature (1985) 313:358-363 describes a chimeric gene in which the small subunit gene is joined to the NPT-II gene (encoding bacterial neomycin phosphotransferase II). The fusion protein encoded in both plasmids described in the paper consists of the 57-amino-acid transit peptide and the first methionine of a mature small subunit polypeptide, a 7-amino-acid linker fragment, and NPT-II lacking the first methionine. Apparently this particular fusion product was capable of transferring the NPT-II gene into chloroplasts. Similar results were described in Schreier et al., EMBO Journal (1985) 4:25-32, although the construct used there contained the 57-amino-acid transit sequence and an additional 22 amino acids from the mature small subunit gene linked via six artificial codons to the same NPT-II gene described by Van den Broeck.

The procedure of Lubben and Keegstra, supra, employed a truncated soybean heat shock protein (17.5 kDa) with a soybean SSU transit peptide and 13 amino acids of the N-terminus of the pea SSU mature peptide.

In comparison with a construct having the pea SSU mature peptide in place of the heat shock protein, the translocation of the heat shock protein was substantially less efficient.

When the procedure of Van den Broeck et al. was attempted with the aroA gene, the gene product was not transported into the plant chloroplasts. Thus, the general indications in Van den Broeck et al. and Schreier et al. that the small subunit transmit protein can be used generally to transport peptides to chloroplasts are seen to be inaccurate. There appears to be a requirement for fusion of the aroA gene to a segment of the small subunit gene that includes the transit peptide sequence, the processing sequence, and the post-processing amino-terminal portion of the mature small subunit peptide or other polar sequence. A functional mutant may be employed in which one or more, preferably one to three, amino acids have been replaced by other amino acids, preferably conservatively, to link the transit peptide gene to the aroA gene.

In order to have expression, it will be necessary to have transcriptional and translational regulatory signals. Conveniently, the chimeric gene may be prepared having its own initiation codon and stop codons, so that expression will be initiated and terminated at the appropriate codons. In addition, a transcriptional initiation region, including the RNA polymerase binding site and transcription initiation site, and terminator region will be provided 5' and 3' to the chimeric gene, respectively.

Conveniently, the transcription initiation region may be the native promoter region associated with the transit peptide. However, in many instances, the native transit peptide transcription initiation region may not be acceptable because it does not provide for the desired degree of transcription or constitutive transcription is desired, where the leader peptide

transcription is inducible, or vice versa. Therefore, the native transcription initiation region may be substituted by a different region or the region upstream from the RNA polymerase binding site and transcription initiation may be substituted to provide for inducible transcription.

The transcription initiation regions which may be employed may be derived from various plant gene promoter regions, including such plant genes as small subunit (SSU) carboxylase, acyl carrier protein (ACP) and Elongation Factor 1, or may be bacterial plasmid gene promoters which are functional in plants, such as the opine synthase promoters of the Ti- or Ri-plasmids of Agrobacterium or viral promoters. These promoters include the promoters associated with opine synthesis, e.g., octopine synthase, nopaline synthase, mannopine synthase, agropine synthase, etc. Viral promoters include the cauliflower mosaic virus 35S promoter and region VI promoter.

In much the same manner as the building up of the chimeric gene, an expression cassette can be developed where the chimeric gene is flanked by transcriptional initiation and termination regions to provide for the desired transcriptional regulation. In some instances, and with increasing frequency, expression constructs are developed, where the transcriptional initiation and termination regulatory regions are provided separated by a polylinker having a plurality of restriction sites, where by insertion of or substitution of a small fragment of the linker with the chimeric gene, the chimeric gene may be ligated so as to be under the regulatory control of the expression cassette. The expression cassette is normally carried on a vector capable of replication and selection in one or more hosts, particularly including prokaryotic hosts.

Depending upon the manner in which the expression cassette is introduced into a plant cell, the expression cassette may be initially joined to other DNA regulatory regions. Commonly, the expression cassette will be joined to a replication system functional in prokaryotes, particularly E. coli, so as to allow for cloning of the expression cassette for isolation, sequencing, analysis, and the like. Included with the replication system will usually be one or more markers which may allow for selection in the host, the markers usually involving biocide resistance, e.g., antibiotic resistance; heavy metal resistance; toxin resistance; complementation; providing prototrophy to an auxotrophic host; immunity; etc.

Where the DNA will be microinjected into the host cell, a marker will usually be desirable which allows for selection of those cells in which the injected DNA has become integrated and functional. Thus, markers will be selected which can be detected in a plant host.

Alternatively, one may use a tumor-inducing plasmid, such as a Ti- or Ri-plasmid, which may be armed (causing gall formation) or disarmed (not causing gall formation). See Hoekema, Nature (1983) 303:179-180; EPA No. 116718 and WO 86/03776. Various constructions can be prepared involving T-DNA where the expression cassette of interest may have a single T-DNA border, the right border, or be bordered on both the right and left sides by T-DNA borders. The borders will involve at least about 50bp of the T-DNA, usually at least about 100bp. If desired, one or more of the T-DNA structural genes with their expression systems may be included, so that integrating into a plant host will result in expression of an opine, which may be used as a marker for the presence of T-DNA integration. The particular manner in which the expression cassette is introduced into the nucleus of the host is not critical

to this invention, so long as the expression cassette is present and capable of functioning to provide the desired product.

The expression cassette included in a prokaryotic vector having one or both T-DNA borders may be transformed into A. tumefaciens or A. rhizogenes containing a Ti- or Ri-plasmid respectively, for example, by the technique described by Zambryski, et al., in Genetic Engineering, Principles and Methods, Vol. 6 (eds. Setlow and Hollaender) 253-278 (Plenum, N.Y., 1984); Binary Vector application Serial No. 834,161; and A. Hoekema, The Binary Plant Vector System (1985), (Offsetdrukkerij Kanters, B.V. Alblessardum). The expression cassette becomes integrated into the tumor-inducing plasmid in the T-DNA, and the bacteria containing the cassette may then be used for infecting plant cells or tissue.

Various techniques exist for determining whether the desired product is present in the plant cell integrated in the genome and is being transcribed. Techniques such as the Northern blot can be employed for detecting messenger RNA which codes for the fused peptide. In addition, the presence of expression can be detected in a variety of ways. Where the expression product provides a detectable phenotype, such as a novel phenotype or enhancement of an endogenous phenotype, the expression of the desired product may be determined by detecting the phenotype. Where a detectable phenotype is not available, antibodies specific for the mature product may be employed. The chloroplasts may be isolated in accordance with conventional ways, disrupted and the Western or other technique employed for identifying the presence of the desired product.

After transformation, the cell tissue (e.g., protoplasts, explants, or cotyledons) are transferred to a regeneration medium for formation of a callus.

The regeneration media will usually contain a bactericide, e.g., carbenicillin (500mg/L), and may contain a selective reagent for selecting transformed cells. For example, with the kanamycin resistance gene (APH3'II), kanamycin will be added to at least about 30mg/L and usually not more than about 500mg/L, preferably from about 50 to 100mg/L, in the selective medium. The regeneration medium includes an appropriate salt source, such as Murashige-Skoog salts medium, a carbon source, e.g., sucrose, with appropriate other additives, such as hormones, e.g., zeatin, etc., at about 0.75-2.25mg/L, myoinositol at about 50-200mg/L, etc. Also, a vitamin supplement may be added, e.g., Nitsch vitamins, at about 0.5 to 1.5ml/L of 1000x stock (usually 1.0ml/L), as is conventional in regeneration media. The 1000x stock of Nitsch vitamins contains in a 100ml final volume: 50mg thiamine HCl, 200mg glycine, 50mg nicotinic acid, 50mg pyridoxine HCl, 50mg folic acid, 5ml biotin and water to volume. The carbon source will be present in from 10 to 30g/L. Conveniently, the regeneration medium contains about 0.5 to 1.0% agar, with the regeneration medium being buffered at about pH 6 ± 0.5.

In 2 to 3 weeks shoots normally develop. When the shoots are approximately 1 to 2 cm, they are excised at the base and transferred to a rooting medium, which may be the same medium on which the seedlings were grown, with carbenicillin and kanamycin sulfate added.

Once the modified plants have been obtained, they can be grown to sufficient maturity to provide sufficient material to determine whether the desired product is still being produced in all or a portion of the plant cells. After expression of the desired product has been demonstrated in the plant, the plants may be bred to seed, used for crossing with other plants, as appropriate, to produce hybrids having the desired

expression capability of the expression cassette or, as appropriate, produce various propagules for further propagation.

The chloroplasts are extremely important cell organelles. Therefore, delivery of proteins to the chloroplasts may allow for the engineering of herbicide resistance, alteration of amino acid metabolism, alteration of photosynthesis, fatty acid metabolism, and other important metabolic functions.

Of particular interest is a construction employing the small subunit leader peptide of 1,5-ribulose-1,5-bisphosphate carboxylase, more particularly, the small subunit leader peptide from soybean (Berry-Lowe et al., J. Mol. Appl. Genet. (1982) 1:43). The leader peptide from this gene may be joined to an aroA gene, particularly an aroA gene expressing an enzyme resistant to inhibition by an enzyme inhibitor for the wild-type enzyme, as is described above.

In conjunction with the chimeric plasmid may be used transcriptional regulation systems, such as those associated with T-DNA, e.g., octopine, nopaline, mannopine or agropine synthases, or the small subunit transcriptional regulation system. Usually, the construction will involve the terminator region of the leader peptide or transcription initiation region, which can be introduced downstream from the structural gene by any convenient means. Usually, the terminator region will be from about 50bp to not more than about 1000bp from the stop codon(s).

The subject procedure results in the production of novel proteins. From one to 20 amino acids of the N-terminus of the mature product may be substituted by the polar region or the N-terminus extended by the polar region. These proteins are distinguished by having an unnatural N-terminus coded for by the hydrophilic region. Thus, as exemplary, is an aroA gene comprising at its N-terminus from 8 to 20 heterologous

amino acids. With the SSU subunit, the protein will comprise 14 to 18, preferably 16 amino acids of the mature SSU protein. Similarly, other of the chloroplast endogenous proteins may be similarly characterized by lacking their N-termini, either as a result of replacement of the natural N-termini with the polar region or by being fused to the polar region to provide an extension of the wild-type N-terminus.

The following examples are offered by way of illustration and not by way of limitation.

## EXPERIMENTAL

### Strains and plasmids

Three RuBP SSU clones were used. pSRS2.1 is a genomic soybean clone (Berry-Lowe et al. (1982) supra). TSSU3-8 and SSU3-2 are genomic tobacco clones (O'Neal et al. Nucl. Acid. Res. (1987) in press). pSS15 is a pea cDNA clone (Coruzzi et al., J. Biol. Chem. (1983) 258:1399). The transit peptide regions of the soybean and tobacco SSU are 75% homologous. The first was used for in vitro uptake studies, the latter for in vivo plant expression.

DNA manipulations were carried out according to established procedures (Maniatis et al. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). Manufacturers specifications were followed where applicable.

### Construction of pCGN565 and pCGN566

pUC12 ($Cm^R$) and pUC13 ($Cm^R$) (Ken Buckley Ph.D. thesis, U.C. San Diego) were each digested with EcoRI and HindIII and polylinkers from pUC18 and pUC19 were inserted respectively into the linearized pUC12 and pUC13 to give pCGN565 and pCGN566 respectively. Each of the plasmids carry a chloramphenicol resistance marker.

Construction of Plasmids Containing the Amino Terminal Portion of RUBPCssu (ribulose-1,5-bisphosphate carboxylase small subunit)

pSRS2.1 (Berry-Lowe et al., supra) containing an EcoRI fragment which includes 5'-untranslated region and the 5'-coding region of RUBPCssu was digested with EcoRI to provide the fragment and the EcoRI fragment inserted into the EcoRI site of pCGN565 and pCGN566, respectively, to provide pCGN330 and pCGN331, respectively. pCGN331 was then further digested with HindIII which removes most of the 5'-untranslated region to provide the amino terminal portion of the RUBPCssu gene carried on a HindIII-EcoRI fragment.

Construction of M13-65B48 (aroA) Clone With Eight Base Pairs Before the ATG Site of the Mutated aroA Salmonella Gene

pPMG11 (Comai et al., Science (1983) 221:370) contains a mutated aroA gene resistant to glyphosate inhibition as a SalI-HindIII fragment. pPMG11 was digested with SalI and BglII and inserted into BamHI-SalI digested pACYC184 (Chang and Cohen, J. Bacteriol. (1978) 134:1141) to provide pPMG17, with the aroA fragment. pUC9 (Vieira and Messing, Molecular Cloning (1982) Cold Spring Harbor Lab Manual) was partially digested with AccI and pPMG17 partially digested with HpaII to provide an about 1.6kbp fragment which was inserted into the AccI site of pUC9 to provide pPMG31. pPMG31 was digested with HindIII, resected about 150bp with Bal31, BamHI linkers ligated to the linear DNA and the DNA digested with BamHI and SalI. The resulting fragment carried the aroA gene with 8bp of the 5'-untranslated region (TGAGTTTC-sense strand). The fragment was inserted into M13mp9 (Messing and Vieira, Gene (1982) 19:269) which had been digested with BamHI and SalI to yield the phage M13-65B48 containing the aroA gene as a BamHI-SalI fragment.

Amino terminal modification of aroA

M13-65B48 was digested with NarI and PstI, generating an aroA fragment lacking the DNA region 5' of the NarI site, corresponding to the amino terminal protion of the gene. This fragment was cloned in pUC8 cut with AccI and PstI, resulting in pPMG58. This plasmid encodes a lac-aroA fusion protein in which 11 amino acids of the pUC8 lac alpha gene replace the 14 amino acids of the N-terminus of the aroA gene. pPMG34 has been described previously (Stalker et al., J. Biol. Chem. (1985) 260:4724-4728). The BamHI to SalI fragment containing the aroA gene was cloned in pUC9, resulting in pPMG34.1. Starting from the BamHI site, the 5'-untranslated region of the aroA gene was recessed by treatments with T4 DNA polymerase in the presence of a single nucleoside triphosphate, followed by mung bean nuclease digestion. After each T4 DNA Polymerase and mung bean treatment, the fragment was subcloned as a blunt to SalI insert into a pUC vector, and transformants were screened for the expected product. After two of these steps a clone into a pUC18 was isolated in which the first nucleotide of the Met codon was part of an XbaI site (...tctagatg). This plasmid, called pPMG34.2, was cut with XbaI, treated with T4 polymerase in the presence of dGTP, followed by mung bean nuclease, and T4 DNA ligase. The DNA was transformed into E. coli LC3, an aroA mutant, (Comai et al., Nature (1983) 317:741-744) and transformants were selected on minimal medium. Several plasmids complementing the aroA phenotype were characterized. Plasmid pPMG34.3 was found to carry an aroA/lac fusion in which the first nucleotide of the aroA Met codon had been deleted.

Construction of SSU-aroA Fusion 1

M13-65B48 was digested with BamHI and SalI and the resulting BamHI-SalI fragment inserted into the BamHI-SalI site of pCGN565. The resulting plasmid, pPMG63, was first linearized with XmaI, followed by digestion with mung bean nuclease to remove the overhangs, and the resulting fragment ligated to SphI linkers, followed by digestion with SphI and SalI.

pCGN566 was completely digested at the unique SphI and SalI sites and the SphI-SalI fragment containing the aroA gene inserted into the site to provide pPMG64 which contains the chloramphenicol resistance gene.

pPMG64 was digested with SphI and a 0.6kb pCGN330 SphI fragment inserted into the SphI site so as to provide the RUBPCssu leader peptide in the proper orientation with the aroA gene. The SphI site is proximal to the 3'-terminus of exon 1 of the RUBPCssu gene. The SphI and BamHI sequences are separated by a single base pair and provide for the RUBPCssu leader peptide and processing signal in proper reading frame with the aroA gene. The sequence is as follows:

| | | | | | | |
|---|---|---|---|---|---|---|
| . . .ACAAT | GCATGC | C | GGATCC | CG | TGACTTTC | ATGGAA. . . |
| | SphI | | BamHI | | 5'-UT aroA region | aroA coding region |

RUBPCss coding region | linker region |

The resulting plasmid pPMG70 was digested with HindIII to remove the upstream untranslated (UT) region of the small subunit to provide plasmid pPMG72. Plasmid pPMG72 was treated with HindIII and EcoRI, and the SSU-aroA promoter/gene fragment was cloned in pSP64 (Melton et al., Nucleic Acid Research (1986) 12:7035) which had been digested with HindIII and EcoRI providing plasmid pCGN1068. The vector pSP64 allows in vitro transcription of cloned DNA.

Construction of SSU-aroA Fusion 3

This fusion was planned in such a way that 24 amino acids of the mature SSU peptide are present between the small subunit transit peptide and the aroA sequence. Plasmid pCGN1068 was cut with XbaI, the resulting stoppered ends were "filled" with the large EcoRI DNA polymerase fragment (Klenow fragment), the plasmid was cut again with SmaI and ligated. This caused the loss of a BamHI site present between the XboI and SmeI site. The resulting plasmid pCGN1075 was cut with EcoRI and HindIII releasing the SSU-aroA chimeric gene (Fusion 1). This gene was cloned in Bluescribe-M13 vector (Vector Cloning Systems, San Diego) cut with EcoRI and HindIII resulting in pCGN1076. Plasmid pCGN1076 was cut with SphI and BamHI and a 70bp SphI-Seu3A fragment from a pea SSU cDNA clone was ligated to pCGN1076 resulting in pCGN1077. Such 70bp SphI-Seu3A fragment was isolated by digesting plasmid pSS15 (G. Coruzzi et al., J. Biol. Chem. (1983) 258:1399) with Seu3A and SphI, separating the resulting fragments by agarose gel electrophoresis and electroelution from the gel. The chimeric gene in pCGN1077 consisted of the transit peptide of the soybean SSU, part of the mature pea SSU (24 amino acids) and the aroA gene. It was called Fusion 3. By digestion of pCGN1077 with EcoRI and HindIII and subsequent ligation of the chimeric gene to pSP64 (D.A. Melton et al., Nucleic Acid Research (1984) 12:7035) cut with HindIII and EcoRI, plasmid pCGN1086 was constructed. The chimeric SSU-aroA gene in pCGN1086 can be transcribed in vitro by the use of SPG-DNA polymerase. A control plasmid containing aroA only was constructed by digesting pPMG63 with BamHI and SalI and cloning it in pSP64 treated with BamHI and SalI, resulting in pCGN1008

Construction of SSU-aroA Fusion 4

To construct an SSU-aroA fusion similar to fusion 3 but lacking the start codon of the aroA gene, the aroA gene of pPMG34.3 was excised as a BamHI to SalI fragment and gel purified. Plasmid pPMG72 was digested with BamHI to SalI. The fragment containing the vector and the SSU coding region was gel purified and ligated to the previously purified aroA gene. The resulting plasmid contained an SSU-aroA fusion lacking the start Met codon of aroA.

In vitro Synthesis and Uptake of Wild-type aroA, Fusion 1 and Fusion 3

pCGN1008, pCGN1068, and pCGN1086 were linearized with EcoRI. They were then transcribed by the addition of SP6-RNA polymerase and nucleotide precursors according to the manufacturer specifications (Promego-Biotec, Madison, Wisconsin). The RNA was then added to in vitro translation extracts from wheat germ (BRL, Bethesda, MD) and translated in the presence of $S^{35}$-methionine (New England Nuclear). The resulting peptides were analyzed by SDS-polyacrylamide gel electrophoresis. Translation of mRNA from pCGN1008 resulted in the synthesis of a 43kd peptide corresponding in mobility to wild-type aroA product. Translation of mRNA from pCGN1068 and pCGN1086 resulted in peptides of respectively 50 and 53kd. In addition, a 43kd peptide was produced from translation of mRNA from pCGN1068, presumably resulting from spurious ribosomal initiation of translation at the original aroA start codon. Each translation product was incubated with isolated spinach chloroplasts as described by Bartlett et al. (in Edelman et al. (eds) Meth. in Chloroplast Mol. Bio. (1982) pp. 1081 Elsevier). The chloroplasts were then treated with trypsin to digest any peptide outside the chloroplasts. Stromal and membrane fractions were separated as described in the reference above and analyzed by SDS-gel electrophoresis.

Construction of the SSU-aroA fusion generation vector pCGN1096

To generate a large number of fusions by the use of nuclease Bal31 resection a specialized vector was constructed called pCGN1096. The following is the construction scheme: the aroA moiety of pCGN1077 was removed by digestion with SphI and SalI. In its place was cloned the region coding for the mature SSU, as an SphI and PstI, and then excising it with SphI and SalI. The resulting plasmid, pCGN1094, codes for an hybrid SSU having the transit peptide of the soybean clone, and the mature portion of the pea clone and carries SstI and EcoRI sites 3' of the coding region. The HindIII to BamHI region of transposon Tn5 (Jorgensen et al., Mol. Gen. Genet. (1979) 177:65) encoding the kanamycin resistance gene was cloned into the same sites of pBR322 (Bolivar et al., Gene (1977) 2:95) generating pDS7. The Bgl2 site 3' of the kanamycin resistance gene was digested and filled in with the large fragment of E. coli DNA polymerase 1 and deoxynucleotides triphosphate. An SstI linker was ligated into the blunted site generating plasmid pCGN1093. Plasmid pPMG34.3 ws digested with SalI, the site filled in as above and EcoRI linkers were ligated into the site resulting in plasmid pCGN1092. The latter plasmid was digested with SstI and SmaI and into it was ligated the kanamycin resistance gene excised from pCGN1093 with SstI and SmaI giving pCGN1095. The kanamycin and aroA gene were excised as a piece from pCGN1095 by digestion with SstI and EcoRI and inserted into the SstI and EcoRI sites of pCGN 1094 giving pCGN1096. Summarizing, pCGN1096 contains 5' to 3' the following pertinent features: The SSU gene - a polylinker coding for PstI, SalI, SstI, and KpnI - the kanamycin resistance gene - SmaI and BamHI restrition sites - the aroA gene without the original ATG start codon.

Generation of SSU-aroA fusions with pCGN1096

Plasmid pCGN1096 was cut at the SalI site and digested with endonuclease Bal31 for different lengths of time. The resected plasmid was then digested with SmaI and religated. Treatment with Bal31 deleted different amounts of SSU coding region starting from the 3' end of the gene. It also deleted different portions of the kanamycin resistance gene starting from the 5' end of the gene. Digestion with SmaI deleted the remainder of the kanamycin resistance gene and ligation joined the ATG-less aroA gene to a point within the SSU coding region. One fusion out of three should generate a translationally active fusion between the SSU coding region and aroA. These events were selected by transforming the resulting plasmids in E. coli LC3, an aroA mutant (Comai et al., (1983) supra), and selecting for the plasmid by ampicillin resistance and for translationally correct fusions by growth on minimal medium. Complementing plasmids were isolated and characterized by restriction digestion. The sequence of the fusion was determined by Sanger dideoxy sequence analysis using a primer homologous to the amino terminal region of aroA.

Construction of the acp-aroA Fusion Generation Vector pCGN1608

A vector analogous to pCGN1096 was constructed to generate acp-aroA fusions. pCGN1SOL contains a cDNA of spinach ACP-I (Scherer and Knauf, Plant Mol. Bio. (1987) 9:127-134). The ACP gene was excised from pCGN1SOL with NcoI and EcoRI, the NcoI end was filled in with the Klenow fragment of DNA PolI, and cloned into pUC18 cut with SmaI and EcoRI to give pCGN1919. The acp carrying plasmid pCGN1919 was digested with EcoRI and into that site was cloned the kanamycin aroA coding region of pCGN1095, resulting in pCGN1608. The relevant region of this plasmid is listed as follows

from 5' to 3': HindIII, PstI, SalI, BamHI - ACP - EcoRI, SstI - kanamycin resistance - SmaI, BamHI - ATG-less aroA - EcoRI, PstI, SphI, HindIII.

Generation of acp-aroA fusions with pCGN1608

Plasmid pCGN1608 was cut at the SstI site and digested with endonuclease Bal31 for different lengths of time. The resected plasmid was then digested with SmaI and religated. Treatment with Bal31 deleted different amounts of acp coding region starting from the 3' end of the gene. It also deleted different portions of the kanamycin resistance gene starting from the 5' end of the gene. Digestion with SmaI deleted the remainder of the kanamycin resistance gene and ligation joined the ATG-less aroA gene to a point within the acp coding region. One fusion out of three should generate a translationally active fusion between the acp coding region and aroA. These events were selected by transforming the resulting plasmids in E. coli LC3, an aroA mutant, and selecting for the plasmid by ampicillin resistance and for translationally correct fusions by growth on minimal medium. Complementing plasmids were isolated and characterized by restriction digestion. The sequence of the fusion was determined by Sanger dideoxy sequence analysis using a primer homologous to the amino terminal region of aroA.

Testing of Fusion Constructs

The fusions chosen for characterization were cloned into an expression vector as follows. Fusions originating from pCGN1608 were cloned as SalI to EcoRI fragments into pCGN1906 cut with the same enzymes. pCGN1906 is an expression vector having the CoMV35S promoter (nucleotides 7146 to 7546) and the OCS 3' region (12,823 to 11,207) (Barker et al., Plant Mol. Bio. (1983) 2:335) cloned via linkers into pUC backbone. SalI and EcoRI sites are located between the

5' promoter and 3' termination region allowing proper positioning of the gene to be expressed. The resulting plasmids contained the fusion under the control of the CaMV 35S promoter and spliced to the tml polyadenylation signal. Fusions originating from pCGN1608 were cloned as BamHI to EcoRI fragments into the expression vector pCGN1096, cut with BglII and EcoRI. The later carries the CaMV 35S promoter and the ocs 3' polyadenylation region separated by BglII, SalI and EcoRI sites. The resulting plasmids were characterized by restriction endonuclease digestion, purified and electroporated into Nicotiana tobacum protoplasts. After 48 hours incubation the electroporated protoplasts were analyzed for the presence of aroA related proteins by Western blot analysis (Comai et al. (1985) supra).

Delivery of Bacterial EPSP Synthase into Isolated Chloroplasts In Vitro

Fusions 1 and 3 incorporate respectively the first 1 and 24 amino acids of the mature SSU. Fusion 4 differs from fusion 3 in the linker-coded region and in the deletion of the aroA start codon.

The fusion genes in the SP6 transcription vector pSP64 (Melton et al., Nucl. Acids. Res. (1984) 12:7035-7056) were cloned and then transcribed and translated in vitro. Translation of mRNA from pCGN1008 resulted in the synthesis of a 43 kDa peptide corresponding in mobility to wild-type aroA product. Translation of mRNA from pCGN1068 and pCGN1086 resulted in peptides of respectively 50 and 53 kDa. Radiolabelled precursor was then incubated with isolated spinach chloroplasts. After incubation the chloroplasts were washed, treated with trypsin if specified, and membrane and stromatic fraction were separated and analyzed by SDS PAGE. Fusion 1 protein is not efficiently translocated into the chloroplasts, as negligible amounts

are found in either the membrane or stromatic fraction. Fusion 3 is translocated: two polypeptides are found in the stromatic fraction and have MW of 47 and 46 kDa, the expected size for the processed product. A change in mobility of the translocated proteins is visible after trypsin treatment. This could indicate that the transport is not complete and part of the protein is still accessible to the protease. In support of this, small amounts of the 47 kDa species of EPSP synthase are present in the membrane fraction. Partial digestion upon trypsin treatment of the chloroplasts could also be explained by limited access to the stroma by trypsun. Alternatively, the 47 kDa species could be the result of an incomplete transport process and, therefore, be partially exposed to proteolysis.

Delivery of Chimeric Fusion Proteins into Chloroplasts by Transient Expression in Leaf Protoplasts

Fusions incorporating 12, 16, 19, 34, 64 and 92 amino acids of the mature small subunit were isolated and characterized. To facilitate the analysis of delivery these fusions were cloned in an expression vector and electroporated into leaf protoplasts. After 48 hours, the protoplasts were analyzed by Western blot using anti aroA antiserum. Although chloroplasts were not isolated, Western blot unequivocally established that the protein was properly delivered and processed since the difference in size between the precursor and the mature protein is about 7 kDa. In no experiment was a precursor ever detected.

The absence of any detectable precursor both in cases of efficient delivery, or in cases of no delivery, indicates that under transient assay conditions, the half life of this species is very short. It is either delivered successfully or turned over rapidly. The amount of mature small subunit present in the fusion has a profound effect on its delivery

efficiency: the most efficient construct was the one incorporating 16 amino acids of the mature small subunit. If its efficiency is given the arbitrary unit of one, the following are the normalized efficiencies: 12 aa = 0.3, 19 aa = 0.09, 24 aa = 0.07, 1 aa = 0.005, 64 and 92 aa <0.003. In addition, the processing heterogeneity observed with fusions 3 (19) and 4 (34) was not seen with either the 12 or 16 aa fusions.

A construct incorporating 35 amino acids of the mature ACP was successfully delivered into the chloroplast although at relatively low efficiency. A second fusion incorporating 42 amino acids was not delivered at all. Although the pattern of use of ACP as a transit peptide is based on only two fusions, it resembles the one of the small subunit. A peptide region at the amino terminus of the mature protein that has been delivered is needed for optimal efficiency. This peptide has salient characteristics; both in the SSU and in the acp it is predicted by the analysis of Kyte and Doolittle, J. Mol. Biol. (1982) 157:105) to be located on the surface of the protein. Furthermore, it carries a series of positively charged amino acids; in the case of the small subunit they are at positions 9, 10, and 11 (all Lys), in ACP there are Lys residues at position 8, 9, 14, 20, 22, and an Arg residue at position 1. Further evidence toward the importance of positively charged amino acids is given by the data of della Cioppa et al., Biotechnology (1987) 5:579-584, on the amino terminal region of the plant EPSP synthase (delivered to chloroplasts as precursor) where positive amino acids residues are found at positions 1, 11, and 18. Lubben and Keegstra, supra used the SSU transit peptide plus 13 amino acids of the mature SSU fused to a soybean heat shock protein to demonstrate delivery in vitro. However, no information was supplied on the effect of the mature portion of SSU for delivery.

It is evident from the above results that improved efficiency of translocation of proteins, particularly those having hydrophobic N-termini can be achieved by employing polar linking groups between the processing signal and protein of interest. In this manner, the N-terminus of the protein may be selected to retain the protein's wild-type characteristics, while providing for substantially complete translocation. In addition, the linker may provide for enhanced or modified properties of the protein. Substantial flexibility is provided in the design of genes and their expression products, while limiting changes at the N-terminus.

All publications cited herein are indicative of the level of skill of those skilled in the art to which this invention pertains. Each publication is individually herein incorporated by reference to the same extent as if each publication had been individually incorporated by reference at the location cited.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A DNA construct comprising in the direction of transcription:

1) a transcriptional and translational initiation region functional in plants;

2) a first DNA sequence coding for a transit peptide and processing signal;

3) a second DNA sequence of from about 6 to 20 codons coding for a polar region comprising at least two positively charged amino acids or of from 12 to 30 codons encoding the N-terminal region of the mature acp gene;

4) a third DNA sequence encoding a gene of interest from other than a plant source; and

a transcriptional and translational termination region functional in plants.

2. A DNA construct according to Claim 1, wherein said first sequence is the transit peptide of the small subunit of ribulose-1,5-bisphosphate carboxylase.

3. A DNA construct according to Claim 4, wherein said second sequence is the natural sequence joined to said first sequence and is of from about 14 to 18 codons.

4. A DNA construct according to Claim 1, joined at its termini to an armed or disarmed Ti- or Ri-plasmid.

5. A DNA construct according to Claim 1, wherein said first sequence is the transit peptide of the small subunit of ribulose-1,5-bisphosphate carboxylase; and said third sequence is the aroA gene with up to 20 of the first codons replaced by said second sequence.

6. A protein comprising (1) a first peptide region consisting of a transit peptide and processing signal functional in a plant cell; (2) a polar region of from about 6 to 20 amino acids; and (3) a second peptide region consisting of a peptide having a function of a chloroplast protein.

7. A protein according to Claim 6, wherein said polar region is the natural sequence joined to said first peptide region.

8. A protein according to Claim 6, wherein said transit peptide is the transit peptide of the small subunit of ribulose-1,5-bisphosphate carboxylase and said second peptide region is the aroA gene with up to 20 of the N-terminal amino acids replaced by said polar region.

9. A protein comprising (1) a first peptide region consisting of the transit peptide and processing signal of a plant acp gene; (2) from 12 to 30 codons encoding the N-terminus of the mature acp; and (3) a second peptide region consisting of a peptide having a function of a chloroplast protein other than acp.

10. A method for introducing a gene of interest into a chloroplast organelle in a plant, said method comprising:

growing plants having in their genome a DNA sequence comprising:

a DNA construct comprising in the direction of transcription:

1) a transcriptional and translational initiation region functional in plants;

2) a first DNA sequence coding for a transit peptide and processing signal;

3) a second DNA sequence of from about 8 to 20 codons coding for a polar region comprising at least two positively charged amino acids;

4) a third DNA sequence encoding a gene of interest other than a plant gene; and

a transcriptional and translational termination region functional in plants;

whereby said DNA sequence is translated to produce a protein comprising said transit peptide, processing signal, polar region and the translation product of said gene of interest, and said protein is translocated to said chloroplast and said transit peptide is removed.

11. A plant cell comprising a construct according to Claim 1.

# INTERNATIONAL SEARCH REPORT

International Application No PCT/US87/02401

## I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [1]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC(4): C12N 15/00, C12N 5/00, C07K 7/00, C07K 13/00
US CL : 435/172.1,240.1; 530/300,350

## II. FIELDS SEARCHED

Minimum Documentation Searched [4]

| Classification System | Classification Symbols |
|---|---|
| U.S. | 435/68,172.1,172.3,183,91,240.4,317.1,320; 47/58<br>935/9,10,30,47,48,49,56,64,67.14<br>530/370,379,300,350 |

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [5]

CHEMICAL ABSTRACT DATA BASE (1967-1987):BIOSIS DATA BASE (1969-1987)KEYWORDS:TRANSIT PEPTIDE,PROTEIN UPTAKE (BY)CHLOROPLAST, PROTEIN PROCESSING,AROA AND GENE,CHLOROPLAST: SEE ATTACHMENT

## III. DOCUMENTS CONSIDERED TO BE RELEVANT [14]

| Category * | Citation of Document, [16] with indication, where appropriate, of the relevant passages [17] | Relevant to Claim No. [18] |
|---|---|---|
| X<br>Y | Nature, volume 313, 31 January 1985, (London, England), Guido Van den Broeck et al, "Targeting of a foreign protein to chloroplasts by fusion to the transit peptide from the small subunit of ribulose 1,5-bisphosphate carboxylase", pages 558-363. See especially 359,360, 361. | 1,2,4,6,10, 11<br>5,8,9 |
| Y | EMBO Journal, Volume 4, August 1985, (Oxford England), P.H.Schreier et al. "The use of nuclear-encoded sequences to direct the light-regulated synthesis and transport of a foreign protein into plant chloroplasts", pages 25-32. See especially pages 26,27,29,30. | 2,4,10,11 |
| Y,P | Biotechnology, Volume 5, June 1987, (New York, New York, U.S.A.), Guy della-Cioppa et al., "Targeting a herbicide-resistant enzyme from escherichia coli to chloroplasts of higher plants", pages 579-584. See especially pages 580,583. | 5 |

* Special categories of cited documents: [15]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

## IV. CERTIFICATION

| Date of the Actual Completion of the International Search [2] | Date of Mailing of this International Search Report [3] |
|---|---|
| 12 December 1987 | 25 JAN 1988 |
| **International Searching Authority [1]** | **Signature of Authorized Officer [20]** |
| ISA/US | Donald Hague |

Form PCT/ISA/210 (second sheet) (May 1986)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

| Category * | Citation of Document, 16 with indication, where appropriate, of the relevant passages 17 | Relevant to Claim No 18 |
|---|---|---|
| Y | Proceedings Of The National Academy Of Science, U.S.A., Volume 83, August 1983 (Washington, D.C.), Thomas H. Lubben, et al., "Efficient in vitro import of a cytosolic heat-shock protein into pea chloroplasts", pages 5502-5506. See especially pages 5503-5505. | 1,2,3,6,7 |
| Y | Nature, Volume 317, 24 October 1985, (London, England), L. Comai, et al Expression in plants of a mutant aroA gene from Salmonella typhimurium confers tolerance to glyphosate", pages 741-744. See especially pages 744-745. | 5,8 |
| Y,P | Journal Of Cellular Biochemistry February 1987, Supplement XI (Part B, Number F018, (New York), John B. Ohlrogge, et al "Plant acyl carrier proteins". Page 10. | 9 |
| A | European Patent Application 0218571A2, D.M. SHAH ET AL., 15 April 1987, Glyphosate-resistant Plants. | |

**FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET**

**V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [10]**

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers . because they relate to subject matter [12] not required to be searched by this Authority, namely:

2. ☐ Claim numbers . because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [13], specifically:

**VI. X OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [11]**

This International Searching Authority found multiple inventions in this international application as follows:

See Attachment

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application. Telephone Practice

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.

PCT/US87/02401

Attachment To Form PCT/ISA/210, Part II.

HETEROLOGUS PROTEIN, ACYL CARRIER
ACYL CARRIER PROTEIN WITH PLANT

Attachment to Form PCT/ISA/210, Part VI.

I. Claims 1-5, drawn to recombinant DNA constructs are classified in Class 435, subclass 172.1.

II. Claims 6-9, drawn to proteins are classified in Class 530, subclass 370.

III. Claim 10, drawn to growing plants is classified in Class 47, subclass 58.

IV. Claim 11, drawn to a plant cell is classified in Class 435, subclass 240.4.

PCT/US87/02401

Attachment to Form PCT/ISA/210, Part VI.1

Telephone Approval:

$420 payment was approved by Mr. Bertram I. Rowland on December 12, 1987 for search of groups 2-4 and will be charged to Deposit Account NO. 12-1216. Counsel was advised that any protest must be filed no later than 15 days from the date of mailing of the search report (Form 210).

Reasons For Holding Lack Of Unity Of Invention:

The claims of these four groups comprise three products and one method and have the characterictics of four distinct inventive concepts. The products of groups I and II are distinct because the DNA constructs are useful as hybridization probes; that is, they do not have to be used in the synthesis of proteins of group II. The products of groups I and II are distinct from group III, a method of growing plants, because the products of group I may be used as probes and the process of group III produces a product other than those of group II. Finally, group IV, a cultured plant cell, is distinct from the products of group I because the construct would be in the genome and the products of group I are useful as probes.

Time Limit For Filing A Protest

Applicant is hereby given 15 days from the mailing date of this search report to file a protest of the holding of lack of unity of invention. In accordance with PCT Rule 40.2 applicant may protest the holding of lack of unity only with respect to the groups paid for.